# ACUTE LYMPHOCYTIC LEUKEMIA CLASSIFICATION USING ENHANCED MACHINE LEARNING AND DEEP LEARNING ALGORITHMS


By

**SARANYA VIJAYAN**

**(Reg. No. 19PHCSF003)**

**Supervisor**

**Dr. V. RADHA**

**Professor, Department of Computer Science**


**Synopsis submitted to the**

**Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore – 641 043**

**In partial fulfilment of the requirements for the Degree of**

**DOCTOR OF PHILOSOPHY IN COMPUTER SCIENCE**

**OCTOBER 2023**

## 1. INTRODUCTION

Leukemia, popularly known as blood cancer, is the result of the rapid overproduction of abnormal white blood cells. Leukemia occurs when abnormal white blood cells in the bone marrow quickly increase and destroy normal blood cells. It is considered to be the $11^{th}$ top cancer type worldwide (Lin *et al.*, 2021). In order to reduce death related to leukemia, it is important to treat them at the early stage.

Approaches used to treat leukemia depends on two factors. The first is whether it is acute or chronic and the second factor is based on the type of the affected white blood cell. Acute leukemia grows rapidly and becomes severe within a short period, while chronic spreads slowly and takes longer to reach the advanced stage. The classification of leukemia is based on the type of white blood cells involving myeloid or lymphoid (Chennamadhavuni *et al.*, 2022). There are four types of Leukemia based on the speed of progression, namely, Acute Lymphocytic Leukemia (ALL) Acute Myeloid Leukemia (AML) Chronic Lymphocytic Leukemia (CLL) Chronic Myeloid Leukemia (CML). Among the four, ALL is the common type of Leukemia, whose detection and classification are the focal point of this research work.

The ALL refers to a type of disease that affects the blood cells. This disease spreads rapidly in the human body and if left undiagnosed and/or untreated, often results in fatal consequences. It is more common with children and adults aged above 50 (Terwilliger and Abdul-Hay, 2017). It can affect the bone marrow all over the body and can also quickly spread to lymph nodes, liver and spleen. A person with ALL may recover if the disease is identified at an early stage.

To ensure ALL detection in a timely manner, the major diagnosis method used is the inspection of peripheral blood samples, in the form of microscopic blood images, to analyze white blood cells. Malformations identified during analysis are a clear indication of the presence of lymphoblasts (Starza *et al.*, 2019). To identify malformations in white blood cells, generally, a manual inspection is carried out by an expert pathologist. Manual inspection and identification has several drawbacks like being time-consuming, high cost as expert pathologists are expensive and diagnosis accuracy depends on the experience and workload of the expert. To avoid these drawbacks, an automatic system is preferred by pathologists, who use these systems

to help them aid in the correct identification of the disease. Several studies have proved that the usage of automatic systems has increased the accuracy and speed of disease identification (Reem *et al.*, 2022; Sharma *et al.*, 2022). Moreover, automated systems help to avoid or reduce human intervention during diagnosis and are cost-effective.

An automatic ALL classification (ALL-C) system consists of four steps, namely, preprocessing, identification of white blood cells, feature extraction and classification. The preprocessing step focuses on improving the quality of the microscopic images so as to help to improve the performance of the subsequent steps of the automatic system by increasing the interpretability of the image. While blood cell identification step extracts the white blood cells from the preprocessed microscopic image. The third step, feature extraction, extracts important characteristics of the white blood cells identified. This step creates a feature vector constructed using the various features extracted. The final step identifies normal and malignant blood cells using a machine learning classifier trained using the constructed feature vector. In the case of malignant blood cells, this step also classifies the blood cell into three of its types, namely, L1, L2 or L3. These four steps are used by several authors during the development of the automated system (Bukhari *et al.*, 2022; Mustafa *et al.*, 2022). The proposed system is designed by enhancing the working of each step of the automated system and then combining them to further improve the performance of the ALL-C system.

Let the template dataset (denoted as D) have M microscopic images collected from N persons. Let $m_i$ refer to the $m^{th}$ microscopic image from the $i^{th}$ person. Let $\oplus$ be the enhancement operation, $\wp$ be the segmentation operation and $\varsigma$ be the identification operation. The research problem is stated as follows.

*"To design an Automatic ALL Classification System of the form*

$$\Re = DC + EMC$$

*where $\Re$ is the output indicating the classification of ALL using hybrid deep learning and ensemble machine learning classifiers. The classification operation is of the form*

$$\mathfrak{R} = \oplus(m_i) \rightarrow \wp(m_i) \rightarrow \varsigma(m_i)$$

*where $\rightarrow$ denotes the sequential application of operations. The classification output is any one of the pre-defined target label set {Normal, L1, L2, L3} and is to be performed using machine learning and deep learning classifiers."*

To accomplish this problem statement, the primary research objective was set to strengthen the clinical decision support system by designing an automatic system that enhances the operation of each step involved during ALL-C in order to increase the overall accuracy and speed of leukemia classification. The specific objectives designed to meet this primary objective are

- To design and develop preprocessing techniques that enhance the visual quality of the microscopic images and identify the white blood cells in them.
- To design and develop an enhanced ensemble machine learning classifier to improve the process of ALL classification.
- To design and develop a deep learning classifier that is combined with ensemble SVM machine learning classifier.

## 2. METHODOLOGY

In order to meet the above-listed objectives, the research methodology was designed in three phases. Each phase was designed separately to satisfy two points. The first was to improve the performance of the task connected to it, while the second was to integrates these tasks together in a manner that help to improve the overall performance of ALL classification. The phases were integrated using a simple I/O (Input/Output) interface, where the output from Phase I was used as input to Phases II and III. The working of each phase along with the optimization methods used is described in the following sections.

## 3. PHASE I : PREPROCESSING

The first phase, preprocessing, performs two tasks, namely, enhancement and white blood cell identification. The enhancement task consists of algorithms that improve the visibility of white blood cells in the microscopic images, by handling the degradations caused by the presence of noise and non-uniform contrast. The second task identifies the Region of Interest (ROI) from the enhanced image using a segmentation algorithm. Here, the ROI is the white blood cells in the enhanced microscopic images.

### 3.1. Noise Removal

Noise in images are visual distortions caused due to random variations of brightness or color information. The presence of noise in images is considered an undesirable by-product introduced during image acquisition and reduces the performance of the ALL-C system. The noise in microscopic images is handled by an algorithm that combines the advantages of two frequently used transformation-based algorithms, namely, Discrete Wavelet Transformation (DWT) and K-Singular Value Decomposition (K-SVD).

DWT-based denoising algorithm provides a simultaneous spatial and frequency domain information of the image and is capable of deconstructing complex signals into basis signals of finite bandwidth, and then reconstructing them again with very little loss of information (Mohideen *et al.*, 2008). However, these algorithms also introduce blocking and ringing artifacts and may degrade fine details, especially in highly noisy images. K-SVD-based denoising algorithm have high noise-removing capability while preserving fine image details and sharp edges (Chen *et al.*, 2021). However, this method has high time complexity and can work well with gray-scale images only. Careful scrutiny revealed that both methods treat noisy and noise-free pixels in the same manner and a noise free-pixel can be treated as noisy pixel, which may result in a blurred image. Further, both algorithms focus only on denoising and do not correct contrast variations. Another point that may affect the ROI-segmentation algorithm is concerned with the edges.

These issues are solved, in this research work, by proposing a unified algorithm that combines contrast adjustment algorithm with noise removal and edge enhancement algorithms. The contrast variations are corrected using an adaptive

histogram equalization algorithm. The noise present in the microscopic image is removed using a hybrid DWT and K-SVD algorithm. This algorithm beings with DWT coefficients to obtain LL, LH, HL and HH subbands, The LL subband is then divided into edge and non-edge regions using its contrast information. The edge region is enhanced using the sigmoid function, while the noise in the non-edge regions are reduced using K-SVD algorithm. During image enhancement, the unified approach proposed offers advantages as listed below.

i. The algorithm is efficient in maintaining structure and edge details to the maximum extent.

ii. Reduces time complexity considerably, as algorithms focus on specific regions

iii. Do not introduce artifacts or blur as noise-free region is not processed at all.

### 3.2. ROI-Extraction

The second task of preprocessing phase is the extraction of White Blood Cells (WBC) from the enhanced microscopic image. It is considered as one of the most important and challenging task of ALL-C system (Vohra and Prodanov, 2021). The challenges faced are due to the high variations of blood cells in terms of shape, size, edge and position. A microscopic image of blood cells has three colors, namely, blue, red and gray-white, which respectively indicate WBC, red blood cells and background. The main objective of the ROI-extraction algorithm is to separate these three cells and is performed using a segmentation algorithm. Segmentation is defined as a task that partitions an image into disjoint and homogeneous regions based on some characteristic of the image.

In order to extract WBCs, several types of segmentation algorithms are used. Each of these algorithms has its own advantages and disadvantages and when applied to the same microscopic image, produces different segmentation results. Some of these results might be more accurate than others. It is a very challenging job to find the correct segmentation algorithm that can produce stable and accurate segmentation results. The search for a best algorithm that can segment a microscopic image is a very difficult and time-consuming job. To solve this issue, instead of comparing

various algorithms to find a perfect segmentation algorithm, this research work proposes a methodological segmentation design, which attempts to increase the accuracy of segments using two segmentation algorithms. The motivation behind this methodology is that it is possible to obtain benefits from combining the strengths of multiple segmenting algorithms.

The methodology behind the proposed segmentation method involves two steps. The first step enhances the working of two conventional algorithms, whose results are then combined to form a final set of segments in the second step. The two algorithms considered during the design of the proposed algorithm are the watershed algorithm and K-means clustering-based algorithm. Both these algorithms were selected because of their success in segmenting images successfully (Vijay and Bhupendra, 2014).

The watershed algorithm is enhanced through the use of a set of techniques, that when applied sequentially can produce accurate segments in a fast manner. The proposed enhanced watershed algorithm is designed using color intensity, Otsu's threshold algorithm, enhanced watershed segmentation algorithm, region merging algorithm and pruning algorithm. The K-means clustering-based segmentation algorithm is enhanced through the use of an automatic technique to determine the initial seeds using a subtractive clustering algorithm. The algorithm sets K as 3 since there are three types of blood cells in microscopic images. The algorithm is further enhanced through the use of a computation reduction algorithm, which can speed up the process of clustering and thus, segmentation. The results of the two enhanced segmentation algorithms are then combined by first generating a mean segmentation image, using which a distance map is constructed. Using this distance map, a weight for each algorithm is estimated. Finally, a majority voting algorithm is used to determine the best segment.

The above multiple segmentation-based algorithm is further enhanced through the use of a post processing procedure. The main objective of this procedure is to improve the perceptibility and visuality of the combined clustering results. The operations of the post-processing procedure are listed below.

- Edge Enhancement - As described in Phase I of the study.

- Morphological Dilation - To connect separated points in a better manner using a 2 x 2 structuring element

- Hole Filling - The internal holes are filled using hole filling method.

## 4. PHASE II : CLASSIFICATION USING MACHINE LEARNING ALGORITHM

Phase II of the research methodology uses the segmented results to classify the identified white blood cells. The steps involved are, feature engineering and classification. Feature engineering consists of two tasks, namely, feature extraction and feature selection. In the classification step, the feature vector obtained from feature engineering is used to classify a cell as normal or cancerous. If cancerous, then to classify them into their types L1, L2 and L3.

### 4.1. Feature Engineering

In this research work, multiple features are extracted from the segmented image, as the usage of multiple features, in place of a single feature, help to improve the accuracy of the classifier. Five types of features are extracted (Mostafa *et al.*, 2019). They are, texture features (Energy, Entropy, Contrast, Correlation, Homogeneity), shape features (Area, Perimeter, Eccentricity, Elongation, Compactness, Minor Axis, Major Axis, Solidity, Form Factor, Nucleus-Cytoplasm Ratio), color features (Mean, Standard Deviation) and irregularity of the nucleus boundary (Horizontal Direction, Vertical Direction). Thus, a total of 19 sets of features are extracted.

The above-created set of features consists of irrelevant and redundant features, which induce greater computational cost and may lead to overfitting. To solve these two issues, a feature selection algorithm is used. Feature selection algorithms work on the principle that not all features are important during classification and these features can be removed without affecting the performance. Feature selection is the task of selecting features that have the maximum impact on describing the results and dropping rest of the features with little or no effect on the performance of the

classifier. This research work proposes the use of the MRMR (Minimum Redundant Maximum Relevant) feature selection algorithm (Fang *et al.* 2020) to improve classification performance in terms of both accuracy and speed.

### 4.2. Classification

Finally, the final step, classification, is designed as an ensemble classification system. The core idea behind the ensemble classification system is to aggregate a set of learners (known as base classifiers) to obtain a combined classification model, that can produce more accurate results than single classifiers. The basic idea of using ensemble systems is to obtain a more accurate and robust classification by considering multiple views of the same problem. Several researchers have used this idea and proved that multiple classifiers work better than single classifier (Sesmero *et al.*, 2021).

An ensemble classification has three main steps, namely, base classifier generation, ensemble creation and aggregation. Base classifiers in an ensemble system can be created in two manners, namely, homogeneous and heterogeneous. Homogeneous refers to the idea of creating the same type of classifier with different parameters or feature sets, while heterogeneous refers to the idea of using different classifiers trained with the same feature set. In this research work, a homogeneous way of creating base classifiers is used. The classifier selected is the most successful Support Vector Machine (SVM). The bagging subspace feature selection technique is used to create multiple feature sets (100 used in this research work), which are then used to create the ensemble system. The output from each of the base classifier is aggregated using a weighted majority voting algorithm.

In this research work, the above-described ensemble system is enhanced in two manners, as listed below.

(i) Usage of optimization procedure that pre-treat the training feature set
(ii) Usage of base classifier selection methods.

The optimization procedure is used to increase the quality of the training set by selecting only those features that help to improve the classification performance. This refined training set helps to obtain a better trained classifier, which in turn helps

to improve the accuracy of the ensemble system. The training feature set quality is improved using a two-step process. In the first step, a single SVM classifier is used to pre-treat the training feature set by collecting all the features that produced true positive and false positive results to form a new optimized training set. The features that produced wrong classification results are removed. The newly constructed refined training set can be used to create a well-trained ensemble system. This step apart from improving the accuracy can also help to reduce the training time complexity of the ensemble system.

The success of the ensemble system depends on the base classifiers selected for constructing the ensemble system, which is still an open and active research. To further enhance the SVM-based ensemble system, an additional step, called 'selection' is included. The selection step is used to select only the best performing base classifier(s), which helps to increase the classification accuracy and to reduce the time complexity.

The selection process can be done either using a static or dynamic approach (Yang, 2011). The static approach is implemented in the training phase, while the dynamic approach is performed for each test sample during the classification phase. The static approach generally uses a pruning algorithm to select a set of best base classifiers. The idea behind the dynamic selection method is that each base classifier is an expert in distinct regions of the feature space. The method aims to select the most competent classifiers in the local region where the new input feature is located. The dynamic approaches can be grouped as (Britto *et al.*, 2014)

- Dynamic Classifier Selection (DCS) - Selects one single best classifier from the set of base classifiers generated
- Dynamic Ensemble Selection (DES) - Selects a subset of best classifiers from the set of base classifiers generated

This research work, to further enhance the process of classification, proposes two enhanced methods that combine static selection and dynamic selection to improve the performance of the proposed EC system. The two proposed hybrid systems are

(i) EC system using static and dynamic classifier selection method

(ii) EC system using static and dynamic ensemble selection method

The static selection is done using a pruning algorithm that selects optimal classifiers among the base classifiers constructed before the training step. The resultant set of classifiers are then supplied to a dynamic ensemble or dynamic classifier selection method, whose results are reported as the final classification output. The methodology used by the proposed ensemble systems is presented in Figure 1.

In this research work, a static pruning technique is used, as a preprocessing function to reduce optimal candidate classifiers. Static techniques work to construct a subset of base classifiers of fixed size to improve its performance with respect to the full ensemble, removing the rest of the classifiers that do not meet this objective (Margineantu and Dietterich, 1997; Zhang *et al.*, 2006; Munoz *et al.*, 2009). The reason for using a static pruning technique with the proposed enhanced EC system is to produce a smaller-sized base classifier set, which can produce the same advantages of the full ensemble system with added advantages like low time complexity.

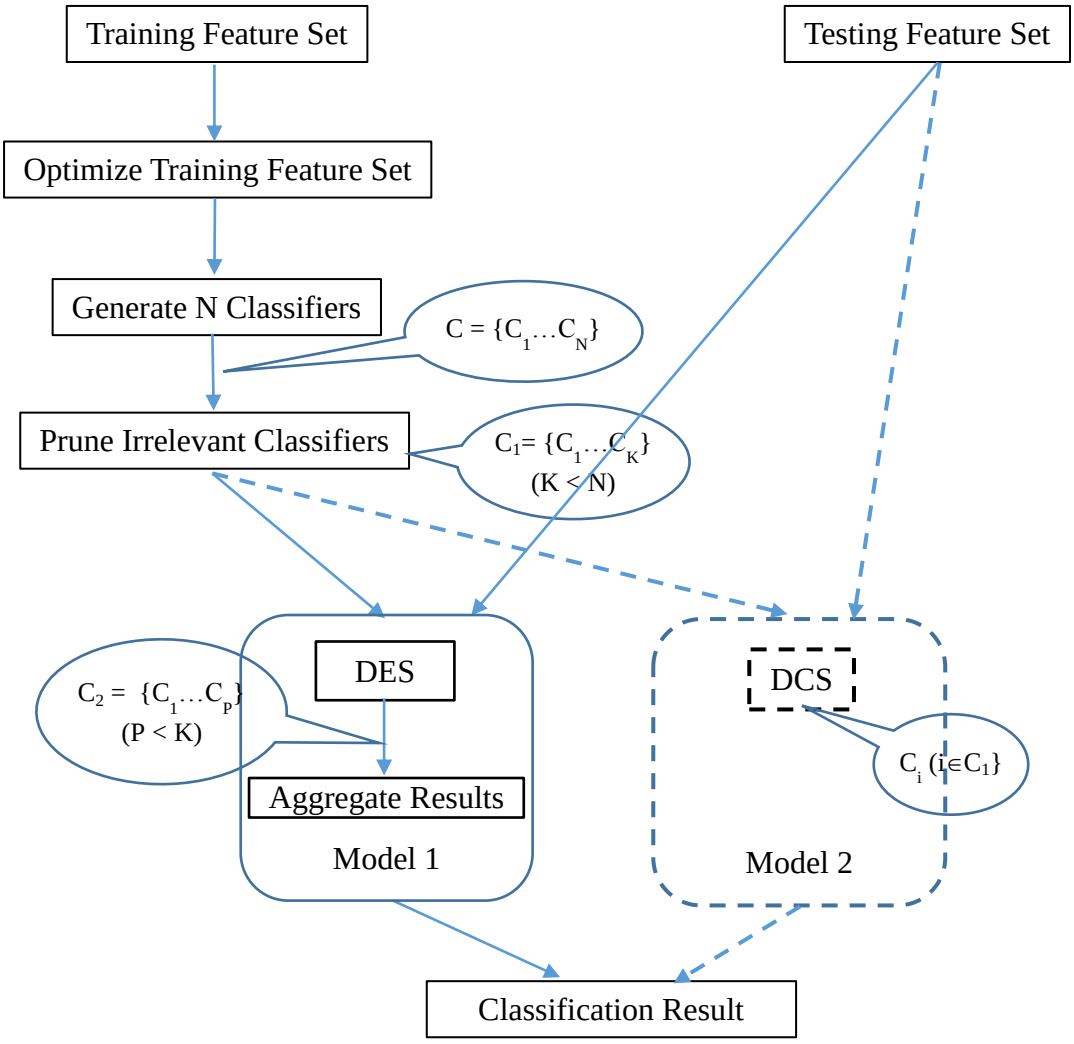

**Figure 1 : Proposed Enhanced Ensemble Classification System**

Let $C_A = \{C_1...C_{50}\}$ be the initial set of base classifier pool created using optimal feature vectors obtained using the Adaboost subspace selection algorithm. The proposed static pruning algorithm, selects a subset of base classifiers from $C_A$, using the following two step-methodology. The first step selects classifiers that reduce the classification error rate. The error rate is defined as the percentage of number of incorrect classifications by total number of classifications. The classifiers are then arranged in ascending order of their error rate and this step returns the top 30 classifiers. Let this set be denoted as $\{C_s\}$.

In the second step, the kappa statistic pruning technique is used to select the final set of optimal classifiers, $\{C_{ss}\}$, which is a subset of $\{C_s\}$. The kappa statistic (Martinez, 2011) is a technique that compares the observed accuracy to an expected accuracy based on the number of instances in each class. This measure is used to get performance information of a classifier when compared to a model that classifies instances at random according to the frequency of each class.

The kappa pruning technique selects a set of classifiers to form an ensemble based on their pairwise diversity using k-statistic. The k coefficients of a pair of classifiers $h_\alpha$ and $h_\beta$ are calculated using the estimated probability that the classifiers coincide in the classification of an instance, $\Theta_1$ and the estimated probability that the classifiers coincide by chance in the classification of an instance, $\Theta_2$. The kappa statistic is then used to measure the agreement between the classifiers. The kappa statistics returns a value between zero and one, with zero indicating poor agreement and 1 indicating perfect agreement (Landis and Koch, 1977). The resulting set of classifiers is denoted as $C_P$.

The aim of dynamic ensemble selection is to find a subset of base classifiers, $C_O$, that can classify a test sample, such that $C_O \in C_P$ and size ($C_O$) < size ($C_P$). In dynamic selection, the classification of test data is done in three steps.

- Step 1 : Region of Competence Identification. Here, a local region surrounding the test data ($Te_i$), is used to estimate the competence level of the base classifier.

- Step 2 : Determine Selection Criteria. This is used to estimate the competence level of the base classifiers. In this research work, the criteria used is the classification accuracy.
- Step 3 : Determine the selection mechanism, that is, DES or DCS.

Region of competence identification is a vital part of the dynamic selection algorithm. In this research work, the local regions are defined using a K-Means clustering-based method. After clustering the training set, the cluster that is nearest to the current test feature is determined. All the nearest clusters are defined as the region of competence for this test data. Using this region, the performance or competence of all the base classifiers are determined for all clusters. Given a new test feature vector, the distance between the test set and centroid of each cluster is estimated. The competence of the base classifiers is then measured based on the features belonging to the nearest clusters. In this research work, the competence measure used is the classification accuracy. That is, all classifiers having the highest accuracy are selected as the most efficient ones, suitable to improve the performance of the EC system. The number of classifiers selected by DCS is one, while for DES, it is set to 15. The best-performing classifier is selected using an automated procedure that uses the classification accuracy as the deciding metric.

## 5. PHASE III : CLASSIFICATION USING DEEP LEARNING ALGORITHM

Currently, there are two methodologies (classical model and deep learning model) followed to perform ALL detection and classification. The classical model denotes the usage of hand-crafted features and machine learning algorithms to perform leukemia detection. Developing such a model was the main focus of Phase II of the research work. As ALL detection is a highly sensitive issue related to the health and life of humans, further algorithms to improve ALL detection were designed in Phase III. Phase III, for this purpose, examines the usage of deep the learning algorithm.

Deep learning (DL) is a transitive subset of artificial intelligence (AI) through machine learning (ML) (Arivuselvam and Sudha, 2022). DL models use neural networks to simulate human-like decision-making. In neural networks, the input

filters through hidden layers of nodes. These nodes each process the input and communicate their results to the next layer of nodes. This repeats until it reaches an output layer, and the machine provides its answer. One major problem with ML models is with the feature extraction process, which is hand-crafted. DL models are capable of focusing on the right features by themselves with very little or no human intervention. It is a class of ML algorithms that exploit many layers of non-linear information processing for supervised or unsupervised feature extraction and transformation for pattern analysis and classification (Sarker, 2021).

To perform early detection of ALL, Phase III proposes a DL classification model based on Convolutional Neural Network (CNN or ConvNet) classifier. This model, called as the base model in this research work, is designed using the AlexNet CNN model based on the transfer learning method, in which deep feature maps were extracted and classified. The CNN classifier was selected because it is one of the most popular types of deep neural networks for image and pattern analysis and recognition (Mayank *et al.*, 2021). A CNN convolves learned features with input data, and uses 2D convolutional layers, making this architecture well suited to processing 2D data, such as images. The CNNs eliminate the need for manual feature extraction and remove the need to identify features used to classify images. CNN works by extracting features directly from images. The relevant features are not pre-trained; they are learned while the network trains on a collection of images. This automated feature extraction makes deep learning models highly accurate for computer vision tasks including ALL classification.

One major issue with CNN is when the size of the training set is small, then the issue of overfitting arises (Xiao *et al.*, 2021). To solve this issue, the training set size is increased through the use of augmentation. Data Augmentation is implemented through the use of image manipulation methods in order to increase the size of the training set. Seven manipulation methods were used. They are, rotation, height shift, width shift, zoom, horizontal flip, vertical flip and shearing. By applying image manipulation methods, the size of the training set was increased and normalized.

The input for the proposed base model AlexNet are the RGB colored images with 227x227 pixel resolution. The base model is designed with 5 convolutional

layers with 3 max polling layers. Each convolutional layer is followed by Rectied Linear Unit (ReLU). For transfer learning, the last 3 layers (fully connected layer, softmax layer and classification layer) of the pre-trained AlexNet are used. Next, the proposed AlexNet architecture was fine-tuned to perform ALL detection and classification, where the data is first classified into normal or abnormal. For this, another fully connected layer with 1024 neurons was added followed by the ReLU layer in order to reduce the feature vector size. All units in this layer are fully connected to 2-class output (normal, abnormal) probability using the softmax function. In the next step, the abnormal category is further classified into L1, L2 and L3 subtype categories. For this, the last fully connected layer is modified from 2 classes to 4 classes output probability, while keeping the rest of the layers the same.

Through experiments, it was evident that the CNN base model outperform the conventional SVM-based model in terms of accuracy and its performance is in par with enhanced ensemble SVM from Phase II. However, CNN has more steps, so, the time needed to run it is longer than SVM. The CNN models are very good at automatically learning the optimal features. However, its performance is not always the best during classification, as the fully connected layers use parameters that have to be fine-tuned manually. To solve this issue, hybrid models that combine CNN and SVM are proposed by several researchers (Kang *et al*., 2018; Liu *et al*., 2018). These systems use CNN for feature extraction and SVM for classification. This research work, moving in the same path, also combines CNN and SVM and includes procedures that can improve its performance.

As mentioned in Phase II of the research methodology, ensemble systems perform better than a single classification system. Therefore, the proposed hybrid system is designed to combine CNN with ensemble SVM. It was noted by several researchers that combining different kernel types can make the ensemble classifier stronger than using a homogeneous SVM trained with different feature vectors but using the same kernel function (Stork *et al.,* 2015; Krishnaveni and Radha, 2021). Using this finding, the ensemble SVM component of the proposed hybrid system is designed to use several SVMs, each differing with the type of kernel function used. The kernel functions used in the proposed system are Linear Kernel, Polynomial Kernel, (RBF), Gaussian Kernel, exponential kernel, Laplacian kernel, Bessel

function kernel, Gaussian Radial Basis Function, ANOVA RBF kernel, Laplace RBF and hyperbolic or Sigmoid Kernel.

Another point from Phase II is that not all base classifiers in the classification system help with classification. For this purpose, the hybrid CNN-ensemble SVM classifier is enhanced to include the selection step described in Phase II. Thus, Phase III of the research methodology proposes three systems as given below:

1. Hybrid CNN-ensemble SVM classifier
2. Hybrid CNN-ensemble SVM using DCS
3. Hybrid CNN-ensemble SVM using DES

All three classifiers listed above works in two stages. In the first stage, the AlexNet CNN extracts the deep features and feeds them to the second stage. The second stage then uses the ensemble SVM classifier to classify the deep feature maps extracted from AlexNet CNNs. The hybrid model proposed replaces the last three fully connected layers of base model by the ensemble SVM classifier. In this stage, dynamic selection methods, DCS and DES, are used, so that only optimal base classifiers are used during classification. Thus, the hybrid model is enhanced by combining CNN with ensemble SVM with DCS/DES method.

The DES method selects five optimal classifiers during the construction of the ensemble system. During aggregation of results, while using ensemble with DES, a weighted ensemble of networks method is used. In this method, the proposed model considers the performance of each SVM using a weight that denotes its contribution to the final classification. The main aim of this method is to enhance classification system performance by using individual SVM's performance to the result of the respective base classifier. The weights are estimated by first calculating the probability of each class and then using it to calculate an evaluation score. The main advantages of the proposed hybrid system are to reduce the chance of overfitting, number of parameters and the time & process complexities.

## 6. EXPERIMENTAL RESULTS

The experiments performed to evaluate the performance of the proposed algorithms in each phase and its cumulative effect on ALL classification was

performed using Acute Lymphoblastic Leukemia Image Database obtained from https://homes.di.unimi.it/scotti/all (Labati *et al.*, 2011). This database has images which were captured using optical laboratory microscope coupled with a Canon PowerShot G5 camera. The images are stored in JPEG format with 24 bit color depth and resolution 257 x 257. There are totally 260 images (130 normal and 130 having leukemia). Additionally 120 more microscopic blood images for typical L1, L2 and L3 type cells were downloaded through Google image search. This additional set of images were used to avoid overtraining. Image manipulations as mentioned earlier were applied to these images and the size of the dataset was increased from 380 to 1760 images (260 normal and 1500 leukemia affected). Out of the 1500 affected images, 666 belonged to L1, 453 belonged to L2 and 381 belonged to L3 types). The experiments were designed in three stages, with each stage focusing on evaluating the algorithms proposed in them.

Stage 1 focused experiments on the evaluation of the preprocessing algorithms (image enhancement and segmentation). The image enhancement algorithms were evaluated using four performance metrics, namely, peak signal to noise ratio, mean structural similarity index, figure of merit and speed of enhancing a single image in seconds. The proposed unified algorithm was compared with three conventional methods, discrete wavelet transformation, K-singular value decomposition and one existing DWT-KSVD algorithm. From the results, it was evident that the proposed unified enhancement algorithm is efficient with all the selected performance metrics. Figure 2 shows the PSNR values obtained by five randomly selected test images, where UCED indicates the proposed unified algorithm. This trend envisaged was the same with all the images in the database.

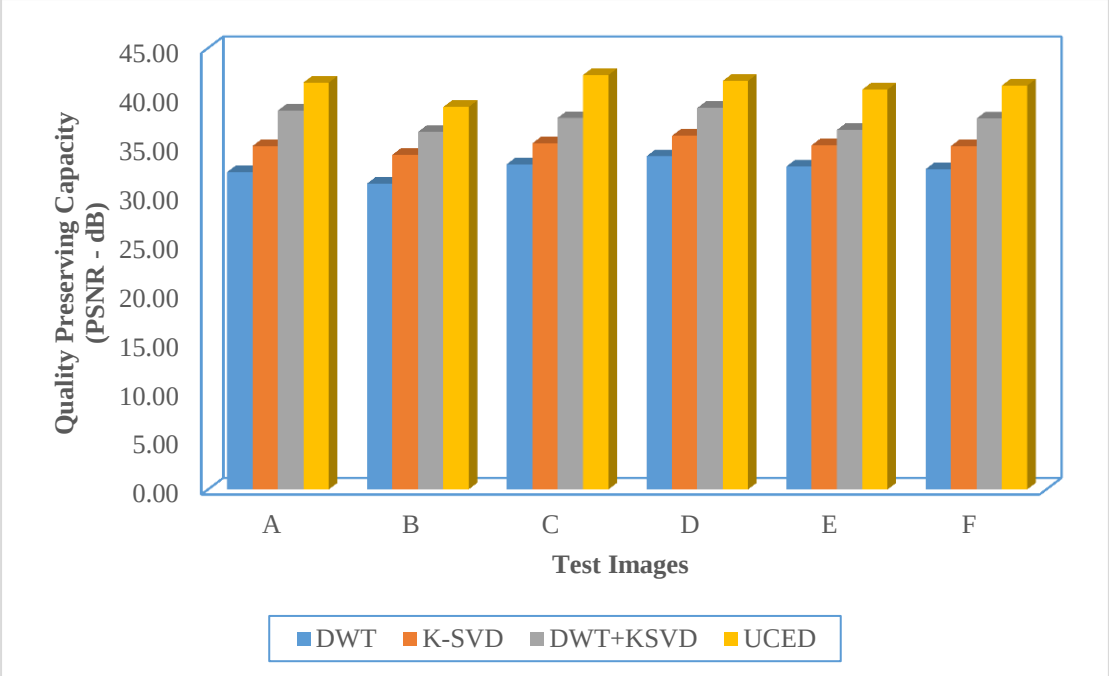


**Figure 2 : PSNR (dB) of Enhancement Algorithms**

From the results obtained while evaluating the segmentation algorithms, it could be seen both the enhanced watershed and enhanced clustering-based segmentation algorithms work better when compared with their respective conventional counterparts. However, the proposed algorithm that combines the results of the two enhanced algorithms finds white blood cells in the most efficient manner.

Stage 2 experiments were used to evaluate the machine learning classification algorithm. Four performance metrics, namely, sensitivity, specificity, accuracy and classification speed were used during evaluation. From Phase II experimental results, it could be understood that the proposed ensemble classifiers, enhanced ensemble SVM using dynamic classifier selection (EDCS-SVM) and enhanced ensemble SVM using dynamic ensemble selection (EDES-SVM), were more powerful, when compared to single SVM (SVM) and conventional ensemble classification systems (ESVM). Moreover, comparison between the proposed classifiers showed that enhanced ensemble SVM using dynamic ensemble selection is slightly more efficient than the enhanced ensemble SVM using dynamic classifier selection. The performance of the classifiers in terms of accuracy is shown in Figure 3.

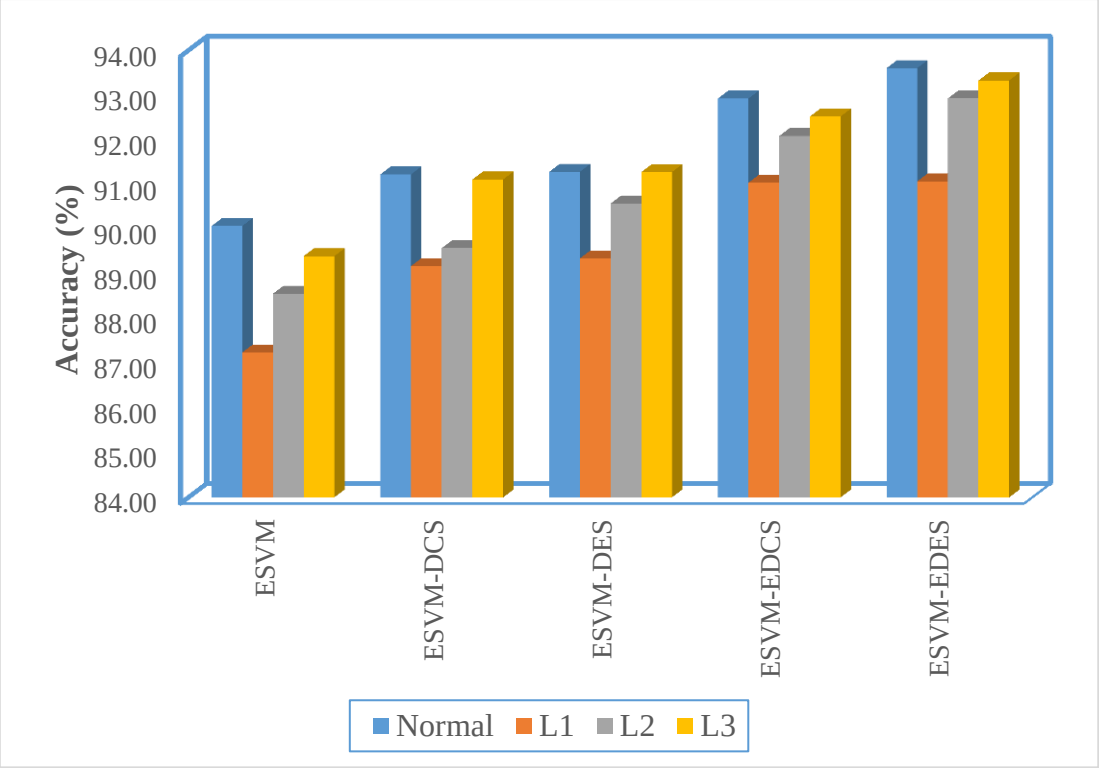


**Figure 3 : Accuracy (%) of Machine Learning Classifiers**

Stage 3 experiments focus on the algorithms proposed in Phase III of the research methodology. This stage of experiments also used sensitivity, specificity, accuracy and speed to evaluate the classifiers. The two proposed algorithm CNN-EDCS-SVM and CNN-EDES-SVM were compared with the conventional SVM, conventional CNN and conventional hybrid CNN-SVM classifiers. Figure 4 shows the accuracy obtained by these classifiers.

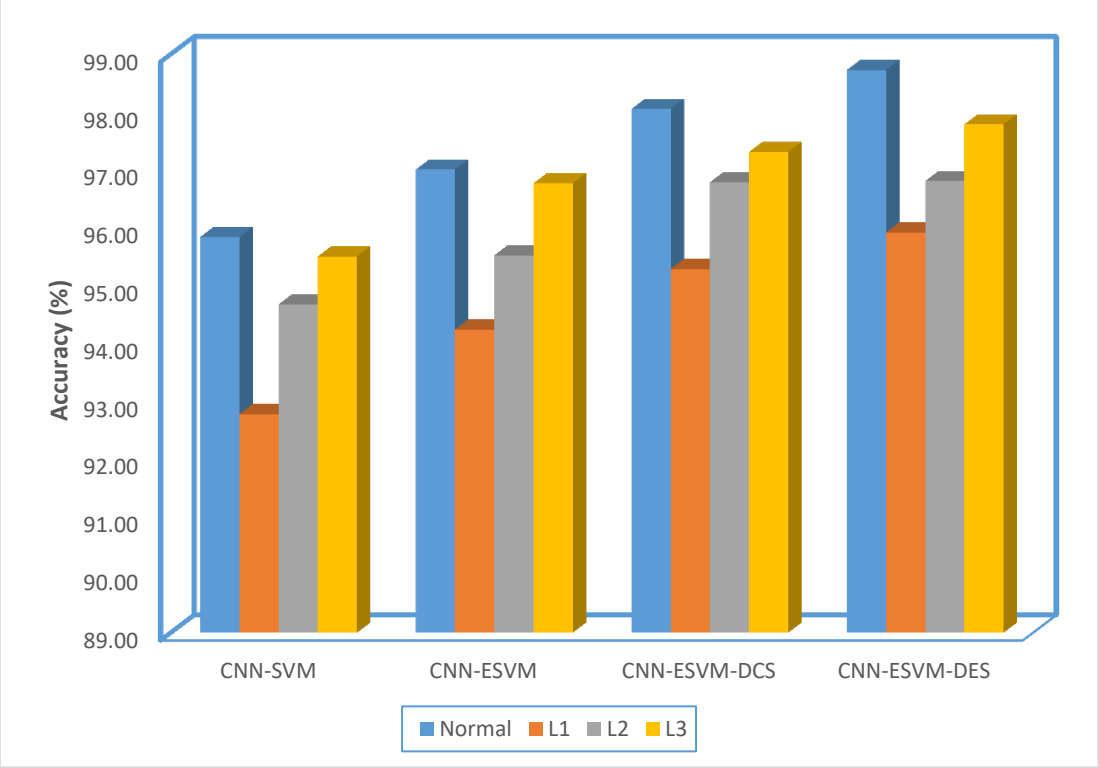


**Figure 4 : Accuracy (%) of Hybrid Classifiers**

From Phase III experimental results, the hybridization of deep learning and machine learning have positive impact on the performance of ALL classification. Comparison of deep and machine learning classifiers showed that CNN outperformed SVM classifier. The proposed hybrid model that used CNN as feature extractor, ensemble SVM with DES for best classifier selection produced maximum advantage during ALL classification. This model produced, on average, a high accuracy of 97.31% accuracy. Thus, from the various results, it could be concluded that the hybrid model that combined CNN with ensemble SVM with dynamic ensemble selection is best suited for ALL classification.

## 7. CONCLUSION

Artificial intelligence algorithms have been used in healthcare industries as powerful analytical and diagnostic tool that can assist physicians with maximum efficiency. ML and DL models can be used to address the various challenges faced during manual diagnosis, differing opinions of experts and time consuming tracking of blood samples. Both models play a key role in the early detection of ALL. This

research work proposed classification models that were enhanced to improve its performance in terms of accuracy during the detection and classification of leukemia.

The research work was divided into three phases. The first phase focused on preprocessing, which works on improving the visual quality of the input microscopic images and identification of the white blood cells. The enhancement of microscopic images was done using a unified approach that combined contrast adjustment, edge enhancement and denoising into a single algorithm. The second task of Phase I is the extraction of white blood cell regions from microscopic images using segmentation algorithm. The proposed segmentation algorithm merged segments produced by watershed and enhanced K-Means-based clustering algorithm. The K-means based clustering algorithm was enhanced by making it parameterless and reducing the number of computations.

Phase II of the research methodology proposed an enhanced SVM based ensemble classifier. The enhancement was achieved by using procedures that reduce the time complexity of ensembling and also to improve its accuracy. The proposed classifier performs classification using two tasks, namely, feature engineering and classification. Feature engineering first extracted four groups of features (texture, shape, color and Irregularity of the nucleus boundary) to form a 19 attribute feature vector. From this feature vector, optimal features were selected using minimum redundant maximum relevant (MRMR) feature selection algorithm. The constructed ensemble model was enhanced through the use of an algorithm that improved the quality of the training data along with dynamic ensemble selection and dynamic classifier selection algorithms. Apart from this, a pruning algorithm that removed irrelevant base classifiers was also proposed.

On the other hand, Phase III of the research methodology, proposed a deep learning model for ALL classification. Initially, a CNN AlexNet base model was designed. This was modified so that the CNN works as a feature extractor, whose results are then fed to SVM to perform classification. This hybrid model was enhanced in two manners. The first was to optimize the training set used and the second is to use ensemble SVM using DES and DCS.

The proposed algorithms were tested vigorously using Acute Lymphoblastic Leukemia Image Database along with images from Google image search. The size of the dataset was increased by using image manipulation methods. Experimental results analyzing algorithms proposed in each phase revealed that the enhancement operations are successful in both improving their respective tasks and improve the performance of ALL classification. A maximum of 98.18% sensitivity, 97.24% specificity and 97.31% accuracy was obtained while using preprocessing methods along with the hybrid model that combined CNN with enhanced SVM-based ensemble model. Thus, it can be concluded that the proposed system meets the research objectives framed and can be used to improve the detection of leukemia disease and can assist pathologist and oncologist to diagnose the leukemia in better way efficiently.

## 8. FUTURE RESEARCH DIRECTIONS

- The proposed system has several tasks. The number of concurrent tasks is limited to only one. Future research can consider parallel processing algorithms to run the tasks simultaneously to save time.
- Future research can also analyse methods to combines features extracted by CNN model with hand crafted feature extraction methods and then using them on the proposed hybrid model to classify data.
- The work can also be extracted by creating hybrid models that combine CNN with other frequently used deep learning models like RNN.